DOCUMENT RESUME

ED 069 690 TM 002 141

AUTHOR	Lewis, Michael; McGurk, Harry
TITLE	The Evaluation of Infant Intelligence: Infant Intelligence Scores--True or False?
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RB-72-32
PUB DATE	Jul 72
NOTE	17p.; A draft
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Infants; *Intelligence Tests; *Intervention; Measurement Instruments; *Predictive Ability (Testing); Program Evaluation; *Research Methodology; Test Reliability; Tests; Test Validity
IDENTIFIERS	Bayley Scales of Infant Development; Peabody Picture Vocabulary Test

ABSTRACT

Infant intelligence from birth until the age of two was measured to determine the usefulness of infant intelligence tests. Twenty infants were tested regularly over the two-year period. Results showed neither simplex nor other long-term patterns of interrelationship among the infant intelligence scores obtained. The study concludes that the concept of general intelligences does not apply to the infancy period and intelligence tests should not, therefore, be used to judge the effectiveness of intervention programs. (DJ)

RESEARCH BULLETIN

RB-72-32

THE EVALUATION OF INFANT INTELLIGENCE:

INFANT INTELLIGENCE SCORES--TRUE OR FALSE?

Michael Lewis

and

Harry McGurk

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. It is automatically superseded upon formal publication of the material.

Educational Testing Service

Princeton, New Jersey

July 1972

The Evaluation of Infant Intelligence:

Infant Intelligence Scores--True or False?

Michael Lewis and Harry McGurk

Educational Testing Service

Abstract

The issue of infant intelligence as a unitary concept was attacked. Using data from three different tests of infant intelligence, it was shown that infant intelligence is neither unitary nor stable over the first two years of life. Implications for intervention programs were discussed in light of the failure to produce meaningful criterion measures of intelligence.

The Evaluation of Infant Intelligence:

Infant Intelligence Scores--True or False?

Michael Lewis and Harry McGurk

The late Sir Cyril Burt once remarked of intelligence, "Of all our mental qualities, it is the most far reaching; fortunately it can be measured with accuracy and ease" (1). Although much progress has been made in the field of psychometrics since Burt's original statement, his early confidence has hardly been justified with respect to the measurement of intelligence during the early stages of human development. In common with many others, Burt espoused a view of intelligence as a finite potential with which the individual was endowed at conception, the manifestations of which increased at a stable rate during the growth process but which was subject neither to qualitative change nor to environmental influence. "...It is inherited, or at least innate, not due to teaching or training; it is intellectual, not emotional or moral, and remains uninfluenced by industry or zeal" (1). It is a sine qua non of such a view that measures of intelligence have high predictive validity from one age to another. Such validity is singularly lacking from every instrument used to assess intelligence during early infancy. For example, Bayley (2), employing an early version of her infant development scales, reported correlations between scores at 1, 2, and 3 months, and scores at 18 to 36 months which ranged between -.04 and .09. Recently, Bayley (3) has concluded, "The findings of these early studies of mental growth of infants have been repeated sufficiently often so that it is now well established that test scores earned in the first year or two have relatively little predictive validity." Stott and Ball (4) and Thomas (5), after extensive reviews covering a wide variety of infant intelligence scales, arrived at essentially similar conclusions.

Despite these acknowledged limitations, infant intelligence scales are widely used in clinical situations in the belief that, although lacking in predictive validity, they provide a valuable aid in assessing the overall health and developmental status of babies at the particular time of testing, relative to other babies of the same age. This procedure is justified only if, in the interpretation of such scores, they are regarded solely as measures of present performance and not as indices of future potential. What this performance may mean is questionable, since it is possible that "superior" performance may be indicative of subsequent poor performance. For example, Bayley shows a negative correlation of -.30 between males' earlier test behavior and IQ at 16-18 years (6). Infant intelligence scales are quite invalid as measures of future potential; the necessity for caution in this respect cannot be overstressed.

Frequently intelligence test scores are used as the criterion measure in the evaluation of the efficiency of infant intervention or enrichment programs. Typically, a sample of subjects from some specified population is exposed to a program of stimulation and interaction beyond the normal experience of the population under study. At various points in the program, intelligence test scores are obtained and compared with those of a sample from the same population, but one not exposed to the enrichment experience. If the scores of the former are higher than those of the latter, the program is evaluated positively; if not, it is evaluated negatively. Two assumptions underlie such procedures, one explicit, the other implicit. Explicitly, it is assumed that, while the limits of intellectual achievement may be genetically determined, mental development is strongly influenced by environmental factors. This is a view which enjoys considerable support, but is not the focus of our interest here. Implicitly, it is assumed that infant intelligence is a general, unitary capacity and that mental development can be enhanced as the result of enriched experience in a few specific areas. Similarly, it is assumed that infant scales are adequate to reflect any improvement that occurs in competence as a consequence of a specific enrichment experience. Data collected in the course of the present longitudinal study of infant affective and cognitive development during the first two years of life made it possible to consider the justification for the latter assumptions.

The present study involved a sample of approximately 20 infants who were seen longitudinally at regular intervals during the first two years (7). There were approximately equal numbers of males and females, and the sample was heterogeneous with respect to social class, although skewed slightly toward the upper levels. The mental scale of the Bayley Scales of Infant Development was administered at 3, 6, 9, 12, 18 and 24 months, as was the Object Permanence Scale from Escalona and Corman's (8) Scales of Sensori-Motor Development. In addition, at two years, language comprehension and production tasks were administered. These were based on a selection of items from the Peabody Picture Vocabulary Test. For the comprehension task, standard Peabody instructions were followed, although a restricted number of items were employed. For the production task subjects were shown individual pictures adopted from the Peabody test and asked, "What is this called?" or "Can you tell me what this is?" Seventeen comprehension and 17 production items were administered to each subject.

Table 1 presents the mean Bayley Mental Development Index (MDI) at each age level, together with standard deviations. It will be noted that at all age levels the mean MDI scores are consistently higher than for Bayley's standardization sample ($\bar{X} = 100$; S.D. = 16). We believe these differences to be a reflection of the relatively high socioeconomic composition of the present sample.

Insert Table 1 about here

Table 2 presents mean scores and standard deviations for the Object Permanence Scale of Sensori-Motor Development. The scale is constructed to reflect the infant's acquisition of the object concept (9); such acquisition is evidenced in the present sample by the regular increase in mean score from one age to the next.

Insert Table 2 about here

Mean scores and standard deviations for the language production and comprehension tasks at two years of age are $\overline{X}$ = 11.53, S.D. 4.66, and $\overline{X}$ = 11.79, S.D. 4.43, respectively.

Intercorrelations between the MDI scores at different ages and between the sensori-motor task at different ages are presented in Table 3.

Insert Table 3 about here

As can be readily seen, of the 30 correlations depicted, only 4 are significant at beyond the .05 level. For the MDI scores, correlations between 3 and 9 months, and between 6 and 24 months reached significance, though in each case the correlation value (.45 and .54) is relatively low and accounts for less than 30 per cent of the variance--relatively useless for predictive purposes. All other MDI correlations are low. Moreover, the data fail to reveal either simplex or other correlational patterns; e.g., 3-month MDI scores predict neither 6-month scores nor 24-month scores (indeed, in the latter instance the correlation is negative). These findings apply across all age levels. Thus, on present evidence, there is neither simplex nor other long-term patterns of interrelationship among infant intelligence scores obtained during the first two years of life (10).

Correlations among scores on the Object Permanence Scale of the Sensori-Motor Development Scales are correspondingly low. Again, only two of them, between 3 and 12 and between 3 and 18 months, reach significance. Both of them account for less than 25 per cent of the variance. Like the MDI scores, there is no clear pattern of interrelationship in the infant's performance on a sensori-motor function. To further stress the point of the lack of interrelationship, other recent work (11) has indicated little or no interrelationship over a variety of sensori-motor scales at any particular age. Thus, although for this sample there is an increase in the mean scale score from one age to another, there is no indication that successful performance at the simpler level will be predictive of an infant's ability to succeed on the more complex items at a later age.

Correlations between MDI and Object Permanence at each age and between language development at 24 months and MDI and Object Permanence at each age are presented in Table 4. The results indicate an interesting developmental

Insert Table 4 about here

pattern of intercorrelations. First, the MDI scales are most related to the object permanence scales of the sensori-motor task in the first six months of life, while the MDI scales are most related to language at 18 and 24 months. This result makes good sense since the early items from the MDI are most related to sensori-motor functions, while the later MDI items are more related to language. Finally, there was no significant relationship between the Object Permanence Scale of sensori-motor functioning and language ability at 24 months. In fact, there are some rather high negative correlations at 9 months.

A number of general conclusions are justified on the basis of these data. Concerning the lack of predictive validity in infant intelligence scales, there is little to add; as in the case of so many other longitudinal studies, present results indicate that there is no reliable relationship between successive measures of infant intelligence during the first two years of life. A similar picture emerges with respect to the measure of sensori-motor development--the Object Permanence Scale--employed in the present study. Although there was a regular increase in mean scores on this scale from one age to the next, and although the majority of subjects showed steady increase in scores over the two-year period (8), high scores at an early age were not predictive of high scores at a later age.

Only at the earlier ages studied was there any significant association between Object Permanence and MDI scores, and this was attributed to the fact that both instruments measure sensori-motor abilities at this period. Beyond 12 months, none of the correlations between the two scales was significant. There was no association between the early MDI scores and the scores on the language tests at _ years; however, there were significant correlations between the MDI scores at 18 and 24 months and the language scores at 24 months. There was no association whatever between scores on the Object Permanence Scale and scores on the language tests. At 2 years, of course, the Bayley test has a considerable verbal loading, whereas the Object Permanence Scale has none.

Overall, these findings cast serious doubt on the applicability of the concept of general intelligence to the infancy period. There is no evidence here to support a view of intelligence as a capacity which unfolds at a steady rate throughout the developmental process and which increases only quantitatively from one age to the next. Rather, present data tend to support the view, advanced by Bayley (3), that at each stage of infant development, intelligence comprises a set of relatively discrete abilities, or factors. During the early developmental period, according to Bayley, these clusters of abilities are relatively age or stage specific, so that there is no necessary continuity between intelligence as defined at one developmental stage and as defined at another. The present data, as well as other information recently reported (11), indicate that even with respect to sensori-motor functioning, there is lack of continuity.

Present data also cast doubt on whether scores gained on infant intelligence scales have any generalizability beyond the particular set of abilities or factors sampled by the items administered at the time of testing. Thus, an infant who showed dramatic gains in tasks involving sensori-motor functioning would not necessarily manifest such gains on tasks involving verbal skills.

The implications of these conclusions for the evaluative policy of infant intervention programs seem clear. Simply stated, infant intelligence scales are quite unsuitable instruments for assessing the effects of specific intervention procedures. This is so primarily because infant intelligence is not a general unitary trait but is, rather, a composite of skills and abilities which do not necessarily covary. Such a view of intelligence is by no means new (12), but it is one which appears to require constant restating in order to counteract a tendency to reify simple, single measures of infant intelligence.

Frequently, the evaluative policy of infant intervention programs has been confused due to a failure to specify clearly the particular set of skills which the program seeks to emphasize and to develop specific criterion tests of those skills. Consider an intervention procedure primarily intended to influence sensori-motor intelligence, for example, the development of object permanence. An appropriate curriculum might involve training subjects in a variety of peek-a-boo and hide-and-seek tasks. It is clear from our data and from the arguments presented above that a standard infant intelligence scale would be the <u>wrong</u> instrument to use in assessing the efficiency of such a program, and that the use of such an instrument is likely to lead to erroneous conclusions concerning the program's efficiency. Even more serious is the possibility that by using the wrong instrument of evaluation over a large number of programs, we would erroneously conclude that intervention in general is ineffective in improving intellectual ability, thus appearing to support the genetic bias that environment is ineffective in modifying intelligence. There are few who would suggest that school children should be administered a standard intelligence test after, say, a course in geography. Yet, such a procedure would be exactly analogous to using an intelligence test to measure the success of attempting to teach the object concept to young infants. Clearly, the success of a geography course is best assessed by tests of geographical knowledge and understanding; by the same token, the success of a program stressing sensori-motor skills is best assessed by specific tests of sensori-motor ability. In both cases, there <u>may in some instances</u> be improvement in intelligence test scores, but such improvement has to be regarded as fortuitous.

It cannot be too strongly emphasized that the success of specific intervention programs geared to improving intellectual functioning must be assessed according to specific criteria related to the content of the program. By focusing attention upon the evaluation criteria, the necessity for careful specification of the program's goals will be emphasized. As argued above, the failure to specify goals has been a contributing factor to the confusion over means of evaluating intervention programs.

The nature and structure of infant intelligence is a complex, and as yet, unsolved problem. In our search for social relevance, we must not be misled into thinking that the validity of our efforts rests solely on the magnitude of the score on an intelligence test of demonstrably limited generality.

References and Notes

1. Burt, C., Jones, E., Miller, E., & Moodie, W. How the mind works. New York: Appleton-Century-Crofts, 1934.

2. Bayley, N. Mental growth during the first three years: A developmental study of sixty-one children by repeated tests. Genetic Psychology Monographs, 1933, 14, 1-92.

3. Bayley, N. Development of mental abilities. In Paul H. Mussen (Ed.) Carmichael's manual of child psychology. New York: Wiley, 1970.

4. Stott, L. H., & Ball, R. S. Infant and preschool mental tests: Review and evaluation. Monographs of the Society for Research in Child Development, 1965, 30, Serial No. 3.

5. Thomas, H. Psychological assessment instruments for use with human infants. Merrill-Palmer Quarterly of Behavioral Development, 1970, 16(2), 179-223.

6. Bayley, N. On the growth of intelligence. American Psychologist, 1955, 10, 805-818.

7. Not all infants were able to complete all tests. The smallest N for which data is reported here is 19. Data from a further 120 infants, seen cross-sectionally at 3, 6, 9, 12, 18, and 24 months (20 infants at each age) were essentially similar to those reported here.

8. Escalona, S. K., & Corman, H. H. The validation of Piaget's hypotheses concerning the development of sensori-motor intelligence: Methodological issues. Paper presented at the Society for Research in Child Development meetings, 1967.

9. Piaget, J. The origins of intelligence in children. New York: International Universities Press, 1952.

10. These results cannot be attributed to low instrument reliability. Bayley has reported split-half reliabilities ranging from .81 to .92 although she also fails to obtain any stability across these early ages.

11. King, W., & Seegmiller, B. Cognitive development from 14-22 months of age in black, male, first-born infants assessed by the Bayley and Hunt-Uzgiris scales. Paper presented at the Society for Research in Child Development meetings, Minneapolis, 1971.

12. Guilford, J. P. Three faces of intellect. American Psychologist, 1959, 14, 469-479.

13. This research was supported by a National Science Foundation grant, No. 28105, and a grant from the Spencer Foundation.

Table 1

Mean and Standard Deviation of

Mental Development Index Scores

	Mean	S.D.
3 Months	101.64	14.9
6 Months	110.05	20.6
9 Months	109.45	13.3
12 Months	113.40	11.6
18 Months	113.63	17.8
24 Months	126.42	18.9

Table 2

Mean and Standard Deviation for Object Permanence Scale of the Escalona and Corman Scales of Sensori-Motor Intelligence

	Mean	S.D.
3 Months	1.10	0.77
6 Months	5.10	1.65
9 Months	8.45	1.90
12 Months	11.80	2.31
18 Months	14.90	1.77
24 Months	15.95	1.39

Table 3

Inter-Age Correlations for the Mental Development Index and Object Permanence Scales[a]

Mental Development Index vs. Mental Development Index

Age (Months)	3	6	9	12	18	24
3		.20	.45*	.06	-.01	-.25
6	-.10		.08	.34	.37	.54*
Object Permanence vs. Object Permanence 9	-.10	00		.40	.13	00
12	.48*	.16	.31		.29	.26
18	.46*	.07	-.13	.32		.36
24	.05	.39	-.07	.08	.05	

*$p < .05$

[a]MDI correlations are in the upper right, while the Object Permanence Scale correlations are in the bottom left of the table.

Table 4

Correlations of the Three Measures of Intellectual Skills

	Months					
	3	6	9	12	18	24
Correlations of MDI with Object Permanence	.24	.60**	.16	.09	.23	.02
Correlations of MDI with Language at 24 Months						
Comprehension	-.19	.40	.10	.22	.42	.49*
Production	-.24	.14	.04	.21	.57**	.48*
Correlations of Object Permanence with Language at 24 Months						
Comprehension	.13	.39	-.28	.17	.38	.21
Production	.21	.26	-.34	-.23	-.15	.31

*$p < .05$

**$p < .01$